DOCUMENT RESUME

ED 240 928 HE 017 034

AUTHOR Brczo, William G.; And Others
TITLE A Study of Test-Wiseness Clues in College and University Teacher-Made Tests with Implications for Academic Assistance Centers. College Reading and Learning Assistance Technical Report 84-01.
PUB DATE [84]
NOTE 17p.; Best copy available.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Instruction; Educational Research; Higher Education; *Multiple Choice Tests; *Teacher Made Tests; *Test Construction; *Test Wiseness

ABSTRACT

A study was conducted to determine to what extent college teacher-made multiple-choice tests contain test-wiseness clues that can be used to identify correct answers. A sample of 43 undergraduate teacher-made examinations was obtained from two colleges and three universities. The sample included midterm and final examinations and quizzes. The tests were written by 36 faculty members, including 9 assistant professors, 23 associate professors, and 4 full professors. A total of 1,220 multiple-choice questions were evaluated based on 10 test-wiseness criteria (Millman, et al., 1965). It was found that 44 percent of the items contained a test-wiseness clue, and 70 percent of these items could be answered correctly by applying a clue. The clue discovered most often was "direct opposites" (i.e, writing an alternative directly opposite to the correct answer). The overall most successful clue was "key word association." It is suggested that college teachers need to consider how to avoid these clues when developing examinations. Implications for college personnel responsible for student academic improvement and faculty development are addressed. Explanations of the 10 test-wiseness criteria are included, along with information on the data analysis procedures. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED240928

College Reading and Learning Assistance
Technical Report 84-01

A Study of Test-Wiseness Clues in College and University Teacher-Made Tests With Implications for Academic Assistance Centers

William G. Brozo
Ronald V. Schmelzer
Hiller A. Spires

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William S Brozo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Georgia State University

RC017034

Counseling services on campus responsible for academic skills training and development have long paid considerable attention to test anxiety. Its debilitating effects on test performance have been well documented (Sarason, 1972; Sarason, 1975; Wine, 1971). Related to the emotional problem of test anxiety is the measurement concern of test reliability. When tests are unreliable, teachers are given spurious estimates of their students' abilities. In the case of the sufferer of test anxiety, his/her true ability may be seriously underestimated due to an inordinate fear of tests. Student anxiety, however, is only one factor that can negatively affect test reliability. Another source of unreliability in testing, recognized by measurement experts, comes in the form of unintentional clues to correct answers inherent in the instrument itself.

Since the introduction of the concept of "test-wiseness" and its effects on test reliability (Thorndike, 1951), efforts have been made by producers of commercially developed standardized tests and conscientious faculty to reduce clues to correct answers within their instruments. Test-wiseness has been defined as a student's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score (Millman, et. al., 1965). Although test-wiseness clues have been a concern of many in higher education, test-wise students still tend to do better than their unsophisticated peers on both standardized and teacher-made examinations (Bajtelsmit, 1977; Brunner, 1976; Flynn & Anderson, 1973; Kirkland & Hollandsworth, 1979; Langer, Wark & Johnson, 1973).

Millman, Bishop and Ebel (1965) developed a set of test-wiseness principles that have been used to help students increase their scores on multiple-choice examinations. These principles provide the test-taker with information on how to select the correct alternative excluding the use of content area knowledge. They have been promoted in various how-to-study books written for students (Brozo & Schmelzer, in press; Morgan & Deese, 1969; Pauk, 1974; Schmelzer, et. al., 1980), as well as in materials specifically designed to teach students test-wiseness (Millman & Pauk, 1969; Sherman & Wildman, 1982; Wark, et. al., 1972; Woodley, 1978).

Based on the knowledge that test-wiseness does exist and that its application seems to have a significant influence on test scores, a study was conducted in an attempt to answer the following question: To what extent do current teacher-made multiple-choice tests at the college/university level contain test-wiseness clues which can be used to identify correct answers?

The study was limited in scope and, therefore, the authors make no sweeping claims about the general ability of all faculty members to construct reliable tests. Nevertheless, it was felt that the research might serve as a model for other academic assistance personnel interested in discovering the extent to which test-wiseness clues are evident in faculty-constructed multiple-choice examinations on their campuses. What is more, if test-wiseness clues to correct answers are found in significant numbers, efforts to bring about more reliable measurements of classroom learning should go in two directions. First, programs should be made available for improving

the faculty's ability to construct multiple-choice tests which are free of clues to correct answers; second, test-wiseness should be taught to unsophisticated test-takers and test anxious students which may help them achieve higher scores.

METHOD

Materials

A sample of forty-three undergraduate teacher-made examinations was obtained from two colleges and three universities. Both colleges (one contributed eight test, the other contributed seven tests) and two of the universities (one contributed ten tests, the other contributed six tests) were located in the Southeastern region of the U.S., with one major university (contributing 12 tests) from the West coast. Included in the sample were mid-term and final examinations, and quizzes. The tests were from the following content areas: business (7); education (5); English (8); geology (3); history (6); health education (6); nursing (6); psychology (2). The tests represented the item writing efforts of 36 faculty members, including nine assistent professors, 23 associate professors and four full professors.

Procedure

A total of 1,220 multiple-choice questions were evaluated according to the following test-wiseness criteria (Millman, et. al., 1965):

1. The longest alternative is usually correct. Operationally, this was defined as an alternative having at least one full line of print more than the other alternatives.

2. The alternative of middle value is usually correct. Operationally this meant that when given a list of alternatives which contained information that could be ordered from large to small, greater to lesser, the alternative between the extremes would most likely be correct.

3. When an alternative contains a key word that is included in the stem, it is usually correct. Operationally this meant that a key word in the stem was readily associated with the same word or a synonym in the alternative.

4. When an alternative contains a specific determiner that allows for exceptions, it is usually correct. Such determiners include the words: often, perhaps, seldom, generally, may, usually, sometimes.

5. When two alternatives are directly opposite in meaning, one is usually correct. To illustrate this criterion, notice the difference between these two alternatives: "decrease as pupils advance through school," "increase as pupils advance through school."

6. If the other alternatives are specific in nature and one is general in nature, than the most general alternative is usually correct. In the following example, alternative "B" is more general than the others. Shakespeare's tragedies: A) deal exclusively with the English monarchs who ruled between 800-1200 (B) deal with the English kings who reigned before the 17th century (C) deal with monarchs of the Tudor line only (D) were written within the first year of his playwriting career.

7. If only one alternative is grammatically consistent with the stem, it is correct. For instance, a stem that reads "Important for

blood regulation are" requires the correct alternative to be plural in nature. If all other alternatives are singular in nature, than the plural alternative must be correct because it agrees grammatically with the stem.

A search was made of three other clues within the 43 tests. The following clues, though not discussed by Millman, were included because of their widespread acceptance by students and teachers of test-wiseness.

8. When in doubt, choose the alternative "all of the above."

9. When most of the test is comprised of four alternative items with some items containing five alternatives, select the fifth alternative.

10. When in doubt, select alternative "C."

Only these clues were considered in this study. Others were excluded because of the difficulty in developing a workable operational definition that could be utilized by the researchers.

Tests were divided among the three researchers and each test item was evaluated according to the test-wiseness criteria above. When there was doubt whether a test item fit the criteria for a particular test-wiseness clue, it was presented to the other researchers and was discarded if one or more of the researchers disagreed.

Data Analysis

Data were analyzed in the following manner:

1. Items that met the criteria for a test-wiseness clue were categorized by the clue and totaled (Clue Related Items). No item was found to possessmore than one clue.

2. Within each category, the number of Clue Related Items which could be answered correctly by using the clue was totaled (Observed Correct).

3. The total of Clue Related Items within each category was multiplied by 25 percent, based on a four alternative item, to determine how many items could be answered correctly by chance. For those items associated with the clue "select the fifth alternative," the total was multiplied by 20 percent (Expected Correct).

4. The difference between the Observed Correct and Expected Correct answers within each category was tested using the normal curve approximation binomial test.

5. For the clue "when in doubt select alternative 'C'," Expected Correct was calculated from the total number of items analyzed in the study (1,220), .25 x 1,169 (four alternative items) and .20 x 51 (five alternative items).

6. The binomial test was computed to determine the difference between the total number of Observed Correct and Expected Correct answers for all clues. The clue "when in doubt select alternative 'C'" was exluded because this clue overlaps all other clues.

RESULTS

The research question was concerned with how often test-wiseness clues appear in college teacher-made multiple-choice tests and to what extent these clues are associated with correct answers. Table 1 shows that out of a total of 1,220 multiple-choice items analyzed, 539 (44%) contained a test-wiseness clue. Three hundred and seventy-

(PLACE TABLE 1 ABOUT HERE)

TABLE 1

Test-Wiseness Clues: Their Frequency
And Probability in a Sample of College Teacher-Made Tests (N=43)

Clue	Clue Related Item	Observed Correct (%)	Expected Correct (%)	z-Score	p Value
1. Longest	54	41 (.76)	13.5 (.25)	8.49	p<.001
2. Middle Value	79	65 (.82)	19.75(.25)	11.62	p<.001
3. Key Word Association	38	38 (1.00)	9.5 (.25)	10.49	p<.001
4. Specific Determiner(True)	52	18 (.35)	13.0 (.25)	1.44	p<.075
5. Direct Opposites	151	115 (.76)	37.75(.25)	14.42	p<.001
6. Most General	45	36 (.81)	10.0 (.25)	3.36	p<.001
7. Grammatical Agreement	0	---	---	---	---
8. All of the Above	69	37 (.54)	17.75(.25)	5.34	p<.001
9. Fifth Alternative	51	25 (.54)	10.2 (.20)	5.00	p<.001
Summary of All Clues	539	375 (.70)	132.2 (.25) (.20)*	24.10	p<.001
10. Alternative "C"	1220	318 (.26)	302.45(.25) (.20)*	1.02	p<.154

*For five alternative items (N = 51)

five (70%) of these items could be answered correctly by applying a clue. Compared with the chance of choosing the correct answer by guessing (132.2 items, or .25 x 488 and .20 x 51), 375 represented a highly significant number of items ($z = 24.10$, $p < .001$).

Table 1 also reveals that all but two of the clues were significantly associated with the correct answer ($p < .001$).

The clue discovered most often was "Direct Opposites." Found in 151 items, one of the two opposite alternatives could be answered correctly in 115 cases. It may be due to the ease with which to write an alternative directly opposite of the correct answer that college teachers use this strategy so often to generate response options for multiple-choice items.

The overall most successful clue was "Key Word Association." Although this clue did not appear often (3% of all 1,220 items), there was a one-to-one correspondence between the clue and the correct answer.

The clue related to grammatical agreement between the stem and one of the alternatives was not found at all. Professors are apparently well aware of this rather obvious clue.

The clue "Alternative 'C'" was of no more help in choosing the correct answer than chance guessing. This suggests that teachers are either making an effort to distribute the correct answer equally among the alternatives, or they are mindful of at least another possible test-wiseness strategy their students might employ.

There was no significant difference between Observed Correct and Expected Correct for the clue "Specific Determiner for True."

On a practical level, however, a student could have obtained five more correct answers by applying the clue instead of guessing.

CONCLUSIONS AND RECOMMENDATIONS

It was observed that most of the test-wiseness clues analyzed in this study were highly significant predictors of the correct response for a given test item. This fact points up the need for college teachers to give more consideration to avoiding these clues when developing their examinations. This is important if teachers desire to avoid penalizing the unsophisticated test-taker as well as increase the chance that the test is a reliable measurement of students' classroom learning.

Due to the limited sample of university and college teacher-made tests in this study, generalization of these findings to all faculty cannot be made. Nonetheless, this research provides some important direction for college personnel responsible for student academic improvement and faculty development.

First, based on the design of this study, a university-specific evaluation of faculty-constructed multiple-choice tests should be made. If test-wiseness clues to correct answers appear in significant numbers, then item writing training workshops should be made available to faculty for improving the reliability of their examinations.

Mehrens and Lehman (1973) characterize a "good" test writer as having a thorough understanding of subject matter and the pupils being tested as well as the ability to use creativity in writing objective questions while avoiding clues to the correct answers.

Other educational measurement specialists (Anderson, 1977; Bloom, et. al., 1971; Gronlund, 1973) have emphasized that writing adequate multiple-choice items requires time, practice, and a checklist of common errors that should be avoided. A checklist should include general questions about multiple-choice test item writing (Anderson, 1977):

Does the stem introduce what is expected of the student?

Is the stem free from irrelevant material?

Are all of the alternatives plausible and homogeneous?

Have you avoided repeating words or phrases in the alternatives?

Have you avoided overlapping alternatives?

Are all alternatives approximately the same length?

If you use the incomplete statement format, does the alternative come at the end of the statement?

The findings of this study suggest that a checklist should also include detailed questions about test-wiseness clues. The following are some examples of questions based on clues reported here.

Is the longest alternative the correct answer?

Is the alternative of middle value the correct answer?

If there are direct opposites among the alternatives, is one the correct answer?

Are there any key words in the stem readily associated with the same word or a synonym in the alternatives?

Clues to the correct answers in multiple-choice tests may never be totally eliminated. Nevertheless, the checklist above represents a start in the development of a more exhaustive checklist that can be used in item writing workshops to improve a college teacher's ability

to write multiple-choice items. Additions to the list will come from the teacher's own experiences with multiple-choice tests as well as from test-writing experts.

The study also points out the continued value of teaching test-wiseness to students who are unsophisticated in its use. This may be particularly true for students not familiar with multiple-choice tests (foreign students, special admission students) in order to avoid the unfair advantage other more sophisticated students may have over them in test taking skills. In addition, teaching test-wiseness clues to test-anxious students may help lower their anxiety. Support for this idea can be found from a number of researchers who brought about a decrease in test anxiety and an improvement in other personality dimensions through academic skills training (Garrison, 1971; Johnson, 1975; Long, 1972; Roth, 1969; Sayles, 1965). Often a student's fear of tests is rooted in a lack of confidence about his/her ability to take tests. This interaction between personality and skills should be a primary concern of college learning center personnel.

According to Anderson, the ideal classroom measurement situation occurs when all test-takers are equally sophisticated and the test writer has designed an instrument as error free as possible. In this case, the chance for reliable measurement is greatly enhanced. College academic assistance personnel are in the unique position of being able to provide programming for faculty and students alike which encourages reliability in classroom testing.

REFERENCES

Anderson, L.W. Multiple choice or multiple guess? University of South Carolina Educational Report, 1977, 20, 1-4.

Bajtelsmit, J.W. Testwiseness and sytematic desentization: programs for increasing adult test-taking skills. Journal of Educational Measurement, 1977, 14, 335-341.

Bloom, B.S., Hastings, J.T., Madaus, G.G. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Brozo, W.G., & Schmelzer, R.V. Setting the pace: a speed, comprehension and study skill program. Columbus, Ohio: Charles Merrill (in press).

Brunner, W.D. Test-wiseness in a college sample: its identification, remediation and influence on achievement. Dissertation Abstracts International, 1976, 36 (9-A), 5936.

Flynn, J.T., & Anderson, B.E. The effects of test item cue sensitivity on I.Q. and achievement test performance. Educational Research Quarterly, 1977, 2, 32-39.

Garrison, C.B. A comparative investigation of behavioral counseling group techniques used to modify study skills, attitudes and achievement on selected high school pupils. Doctoral dissertation, State University of New York at Buffalo, 1971.

Gronlund, N.E. Constructing achievement tests. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Johnson, S.D. Skills therapy: a description of some changes on selected personality variables produced by academic skills training. Doctoral dissertation, University of Minnesota, Minneapolis, 1975.

Kirkland, K., & Hollandsworth, J.G. Test anxiety, study skills and academic performance. Journal of College Student Personnel, 1979, 20, 431-435.

Langer, G., Wark, D., & Johnson, S. Test-wiseness in objective tests. In P.L. Nacke (Ed.), Diversity in Mature Reading: Theory and Research. Twenty-second Yearbook of the National Reading Conference, 1973.

Long, R.W. A study of the relationship between academic achievement and changes in self-concept in a population of adult high school students. Doctoral dissertation, Florida State University, 1972.

Mehrens, W.A., & Lehman, I.J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart, and Winston, 1973.

Millman, J., Bishop, C.H., & Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Millman, J., & Pauk, W. How to take tests. New York: McGraw-Hill, 1969.

Morgan, C.T., & Deese, J. How to study. New York: McGraw-Hill, 1969.

Pauk, W. How to study in college. Boston: Houghton, Mifflin, 1974.

Roth, R.M. The role of self concept in achievement. Journal of Experimental Education, 1969, 27, 265-281.

Schmelzer, R.V., Christen, W., & Browning, W. Reading and study skills. Rehoboth, Mass.: Twin Oaks, 1980.

Sarason, I.G. Experimental approaches to test anxiety: attention and the uses of information. In C. Spielberger (Ed.) Anxiety: current trends in theory and research (Vol. 2). New York: Academic Press, 1972.

Sarason, I.G. Test anxiety and the self-disclosing coping model. Journal of Consulting and Clinical Psychology, 1975, 43, 148-153.

Sayles, D. G. Review of research: skill therapy. In D. Wark (Ed.) College and Adult Reading. Minneapolis, Minnesota: University of Minnesota, North Central Reading Association, 1965.

Sherman, T.M., & Wildman, T.M. Proven strategies for successful test taking. Columbus, Ohio: Charles Merrill, 1982.

Thorndike, E.L. Reliability. In E.F. Lindquist (Ed.) Educational Measurement. Washington, D.C.: American Council on Education, 1951.

Wark, D., Moore, F., Langer, G., Pinchot, J., & Tostenud, D. A script to teach test wiseness. University of Minnesota, Office of International Programs, 1972.

Wine, J.D. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76, 92-104.

Woodley, K.K. Test wiseness: test-taking skills for adults. New York: McGraw-Hill, 1978.